Internet Appendix to

Are Stocks Really Less Volatile in the Long Run?

by

Ľuboš Pástor

and

Robert F. Stambaugh

January 10, 2011
B1. Roadmap

This Appendix is organized as follows. A simple illustration demonstrating the effects of parameter uncertainty on long-horizon predictive variance is provided in Section B2. The Bayesian empirical methodology for System 2 is presented in Section B3. Additional empirical evidence that complements the evidence presented in the paper is reported in Section B4. That section first presents the estimates of predictive regressions for both annual and quarterly data, followed by various robustness results regarding the long-horizon predictive variance of stock returns.

B2. Parameter uncertainty: A simple illustration

In the paper, we compute $\text{Var}(r_{T,T+k}|D_T)$ and its components empirically, incorporating parameter uncertainty via Bayesian posterior distributions. Here we use a simpler setting to illustrate the effects of parameter uncertainty on multiperiod return variance. All the notation that is not defined here is defined in Section 2 of the paper.

The elements of the parameter vector $\phi$ are viewed as random, given that they are unknown to an investor. For the purpose of this illustration, we make several assumptions related to $\phi$ with the objective of simplifying the variance ratio calculations in the presence of parameter uncertainty. Our first assumption is that, given data $D_T$, the elements of $\phi$ are distributed independently of each other. We also make two assumptions about the deviations of the conditional mean from the unconditional mean: given $D_T$, $\mu_T - E_r$ is uncorrelated with $E_r$, and the quantity $z_T \equiv b_T - E_r$ is fixed across $\phi$. Next, let $\rho_{\mu b}$ denote the true unconditional correlation between $\mu_T$ and $b_T$, $\rho_{\mu b} \equiv \text{Corr}(\mu_T, b_T | \phi)$ (this correlation is “true” in that it depends on the true parameter values $\phi$, and it is “unconditional” in that it does not condition on $D_T$). If the observed predictors capture $\mu_T$ perfectly, then $\rho_{\mu b} = 1$; otherwise $\rho_{\mu b} < 1$. We make the “homoskedasticity” assumption that $q_T$ is constant across $D_T$, which implies that

$$q_T = (1 - \rho_{\mu b}^2)\sigma_{\mu}^2 = (1 - \rho_{\mu b}^2)R^2\sigma_r^2,$$

(B1)

where $\sigma_{\mu}^2$ and $\sigma_r^2$ are the true unconditional variances of $\mu_t$ and $r_{t+1}$, respectively (i.e., $\sigma_{\mu}^2 \equiv \text{Var}(\mu_t | \phi)$ and $\sigma_r^2 \equiv \text{Var}(r_{t+1} | \phi)$). The parameter vector is $\phi = [\beta, R^2, \rho_{uw}, E_r, \sigma_r, \rho_{\mu b}]$. Finally, we specify $\tau$ such that

$$\text{Var}(E_r|D_T) = \tau \text{E}(\sigma_r^2|D_T),$$

(B2)

so that the uncertainty about the unconditional mean return $E_r$ is as large as the imprecision in a sample mean computed over a sample of length $1/\tau$. 

1
None of the above simplifying assumptions hold in the Bayesian empirical framework in the paper; they are made for the purpose of this simple illustration only. Given these assumptions, we are able to characterize the variance ratio in a parsimonious fashion that does not depend on the unconditional variance of single-period returns.

The variance of the $k$-period return can be decomposed as follows:

\[
\text{Var}(r_{T,T+k}|D_T) = E \left\{ \text{Var}(r_{T,T+k}|\phi, \mu_T, D_T)|D_T \right\} + E \left\{ \text{E}(r_{T,T+k}|\phi, \mu_T, D_T)|D_T \right\}
\]

\[
= E \left\{ \text{Var}(r_{T,T+k}|\phi, \mu_T)|D_T \right\} + E \left\{ \text{E}(r_{T,T+k}|\phi, \mu_T)|D_T \right\}
\]

\[
= E \left\{ k\sigma^2_r(1 - R^2)[1 + 2\bar{\rho}_{uw}\beta A(k) + \bar{d}^2 B(k)]|D_T \right\}
\]

\[
+ \text{Var} \left\{ kE_r + \frac{1 - \beta^k}{1 - \beta}(\mu_T - E_r)|D_T \right\}
\]

\[
= kE(\sigma^2_r|D_T)E \left\{ (1 - R^2)[1 + 2\bar{\rho}_{uw}\beta A(k) + \bar{d}^2 B(k)]|D_T \right\}
\]

\[
+ k^2 \text{Var}(E_r|D_T) + \text{Var} \left\{ \frac{1 - \beta^k}{1 - \beta}(\mu_T - E_r)|D_T \right\}
\]

\[
= kE(\sigma^2_r|D_T)E \left\{ (1 - R^2)[1 + 2\bar{\rho}_{uw}\beta A(k) + \bar{d}^2 B(k)]|D_T \right\}
\]

\[
+ k^2 \tau E(\sigma^2_r|D_T) + \text{Var} \left\{ \frac{1 - \beta^k}{1 - \beta}(\mu_T - E_r)|D_T \right\} . \quad (B3)
\]

The next to last equality uses the property that $E_r$ is uncorrelated with $\mu_T - E_r$, and the last equality uses (B2). Next observe that

\[
\text{Var} \left\{ \frac{1 - \beta^k}{1 - \beta}(\mu_T - E_r)|D_T \right\} = E \left\{ \text{Var} \left\{ \frac{1 - \beta^k}{1 - \beta}(\mu_T - E_r)|\phi, D_T \right\}|D_T \right\}
\]

\[
+ \text{Var} \left\{ E \left\{ \frac{1 - \beta^k}{1 - \beta}(\mu_T - E_r)|\phi, D_T \right\}|D_T \right\}
\]

\[
= E \left\{ \left( \frac{1 - \beta^k}{1 - \beta} \right)^2 \text{Var}(\mu_T|\phi, D_T)|D_T \right\}
\]

\[
+ \text{Var} \left\{ \frac{1 - \beta^k}{1 - \beta}E(\mu_T - E_r|\phi, D_T)|D_T \right\}
\]

\[
= E \left\{ \left( \frac{1 - \beta^k}{1 - \beta} \right)^2 q_T|D_T \right\}
\]

\[
+ \text{Var} \left\{ \frac{1 - \beta^k}{1 - \beta}(b_T - E_r)|D_T \right\}
\]

\[
= E \left\{ \left( \frac{1 - \beta^k}{1 - \beta} \right)^2 \sigma^2_rR^2(1 - \rho_{\mu b}^2)|D_T \right\}
\]

\[
+ \text{Var} \left\{ \frac{1 - \beta^k}{1 - \beta}z_T|D_T \right\}
\]

2
\[ = \mathbb{E}\left\{ \left( \frac{1 - \beta^k}{1 - \beta} \right)^2 | D_T \right\} \mathbb{E}(\sigma_r^2 | D_T) \mathbb{E} \left\{ R^2 (1 - \rho_{\mu b}^2) | D_T \right\} + z_T^2 \text{Var} \left\{ \frac{1 - \beta^k}{1 - \beta} | D_T \right\}. \]  

(B4)

Substituting the right-hand side of (B4) for the last term in (B3) then gives

\[ \text{Var}(r_{T,T+k}|D_T) = k \mathbb{E}(\sigma_r^2 | D_T) \mathbb{E} \left\{ (1 - R^2)[1 + 2d_\rho_{uw} A(k) + d^2 B(k)] | D_T \right\} + k^2 \tau \mathbb{E}(\sigma_r^2 | D_T) \mathbb{E} \left\{ \left( \frac{1 - \beta^k}{1 - \beta} \right)^2 | D_T \right\} \mathbb{E} \left\{ R^2 (1 - \rho_{\mu b}^2) | D_T \right\} + z_T^2 \text{Var} \left\{ \frac{1 - \beta^k}{1 - \beta} | D_T \right\}. \]  

(B5)

When \( k = 1 \), equation (B5) simplifies to

\[ \text{Var}(r_{T,T+1}|D_T) = \mathbb{E}(\sigma_r^2 | D_T) \left[ 1 + \tau - \mathbb{E}(R^2 | D_T) \mathbb{E}(\rho_{\mu b}^2 | D_T) \right]. \]  

(B6)

Observe that the \( k \)-period variance in (B5) depends on the value of \( z_T^2 \), which enters the last term multiplied by the variance of \( (1 - \beta^k)/(1 - \beta) \). This dependence makes sense: when \( \mu_T \) is estimated to be farther from the unconditional mean \( E_r \), so that the absolute value of \( z_T \) is large, uncertainty about the speed with which \( \mu_T \) reverts to \( E_r \) is more important. To achieve a further algebraic simplification, we evaluate the variance in (B5) by setting \( z_T^2 \) equal to the posterior mean of the true unconditional mean of \( z_T^2 \):

\[ z_T^2 = \mathbb{E}(z_T^2 | \phi) | D_T]. \]  

(B7)

To evaluate the right-hand side of (B7), first note that \( \text{Var}(z_T | \phi) = \text{Var}(b_T | \phi) \), since \( z_T = b_T - E_r \) and \( E_r \) is in \( \phi \). Also observe that \( \mathbb{E}(z_T | \phi) = 0 \), since across samples (\( D_T \)’s), the expected conditional mean \( b_T \) is the unconditional mean \( E_r \). Therefore, \( \mathbb{E}(z_T^2 | \phi) = \text{Var}(z_T | \phi) = \text{Var}(b_T | \phi) \). We thus have

\[ \mathbb{E}[\mathbb{E}(z_T^2 | \phi) | D_T] = \mathbb{E}[\text{Var}(b_T | \phi) | D_T] = \mathbb{E}[\rho_{\mu b}^2 \text{Var}(\mu_T | \phi) | D_T] = \mathbb{E}[\rho_{\mu b}^2 \sigma_r^2 R^2 | D_T] = \mathbb{E}(\rho_{\mu b}^2 | D_T) \mathbb{E}(\sigma_r^2 | D_T) \mathbb{E}(R^2 | D_T). \]  

(B8)

We compute the \( k \)-period variance ratio,

\[ V(k) = \frac{\text{Var}(r_{T,T+k}|D_T)}{k \cdot \text{Var}(r_{T,T+1}|D_T)} \]  

(B9)
where $\text{Var}(r_{T,T+1}|D_T)$ is given in equation (B6) and $\overline{\text{Var}}(r_{T,T+k}|D_T)$ denotes the value of (B5) obtained by substituting the right-hand side of (B8) for $z_T^2$. Note that the variance ratio in equation (B9) does not depend on $\mathbb{E}(\sigma_r^2|D_T)$.

We set $\tau = 1/200$ in equation (B2), so that the uncertainty about the unconditional mean return $E_r$ corresponds to the imprecision in a 200-year sample mean. To specify the uncertainty for the remaining parameters, we choose the probability densities displayed in Figure A1, whose medians are 0.86 for $\beta$, 0.12 for $R^2$, -0.66 for $\rho_{uw}$, and 0.70 for $\rho_{\mu b}$.

Table A1 displays the 20-year variance ratio, $V(20)$, under different specifications of uncertainty about the parameters. In the first row, $\beta$, $R^2$, $\rho_{uw}$, and $E_r$ are held fixed, by setting the first three parameters equal to their medians and by setting $\tau = 0$ in (B2). Successive rows then specify one or more of those parameters as uncertain, by drawing from the densities in Figure A1 (for $\beta$, $R^2$, and $\rho_{uw}$) or setting $\tau = 1/200$ (for $E_r$). For each row, $\rho_{\mu b}$ is either fixed at one of the values 0, 0.70 (its median), and 1, or it is drawn from its density in Figure A1. Note that the return variances are unconditional when $\rho_{\mu b} = 0$ and conditional on full knowledge of $\mu_T$ when $\rho_{\mu b} = 1$.

Table A1 shows that when all parameters are fixed, $V(20) < 1$ at all levels of conditioning (all values of $\rho_{\mu b}$). That is, in the absence of parameter uncertainty, the values in the first row range from 0.95 at the unconditional level to 0.77 when $\mu_T$ is fully known. Thus, this fixed-parameter specification is consistent with mean reversion playing a dominant role, causing the return variance (per period) to be lower at the long horizon. Rows 2 through 5 specify one of the parameters $\beta$, $R^2$, $\rho_{uw}$, and $E_r$ as uncertain. Uncertainty about $\beta$ exerts the strongest effect, raising $V(20)$ by 17% to 26% (depending on $\rho_{\mu b}$), but uncertainty about any one of these parameters raises $V(20)$. In the last row of Table A1, all parameters are uncertain, and the values of $V(20)$ substantially exceed 1, ranging from 1.17 (when $\rho_{\mu b} = 1$) to 1.45 (when $\rho_{\mu b} = 0$). Even though the density for $\rho_{uw}$ in Figure A1 has almost all of its mass below 0, so that mean reversion is almost certainly present, parameter uncertainty causes the long-run variance to exceed the short-run variance.

As discussed in the paper, uncertainty about $E_r$ implies $V(k) \to \infty$ as $k \to \infty$. We can see from Table A1 that uncertainty about $E_r$ contributes nontrivially to $V(20)$, but somewhat less than uncertainty about $\beta$ or $R^2$ and only slightly more than uncertainty about $\rho_{uw}$. With uncertainty about only the latter three parameters, $V(20)$ is still well above 1, especially when $\rho_{\mu b} < 1$. Thus, although uncertainty about $E_r$ must eventually dominate variance at sufficiently long horizons, it does not do so here at the 20-year horizon.

The variance ratios in Table A1 increase as $\rho_{\mu b}$ decreases. In other words, less knowledge about $\mu_T$ makes long-run variance greater relative to short-run variance. We also see that drawing $\rho_{\mu b}$
from its density in Figure A1 produces the same values of $V(20)$ as fixing $\rho_{ib}$ at its median.

**B3. Predictive System 2**

System 2 is implemented in the paper for one asset, but it can be specified for $n$ assets, in which case $r_{t+1}$ and $\pi_{t+1}$ are $n \times 1$ vectors. We begin the analysis with that more general form and then later restrict it to the single-asset setting. With multiple assets, System 2 is given by

$$
\begin{bmatrix}
  r_{t+1} \\
  x_{t+1} \\
  \pi_{t+1}
\end{bmatrix} =
\begin{bmatrix}
a \\
\theta \\
0
\end{bmatrix} +
\begin{bmatrix}
  0 & A_{12} & I \\
  0 & A_{22} & 0 \\
  0 & 0 & A_{33}
\end{bmatrix}
\begin{bmatrix}
r_t \\
x_t \\
\pi_t
\end{bmatrix} +
\begin{bmatrix}
  u_{t+1} \\
v_{t+1} \\
\eta_{t+1}
\end{bmatrix},
\tag{B10}
$$

which can also be written as

$$
\begin{bmatrix}
r_{t+1} - E_r \\
x_{t+1} - E_x \\
\pi_{t+1} - E_\pi
\end{bmatrix} =
\begin{bmatrix}
  0 & A_{12} & I \\
  0 & A_{22} & 0 \\
  0 & 0 & A_{33}
\end{bmatrix}
\begin{bmatrix}
r_t - E_r \\
x_t - E_x \\
\pi_t - E_\pi
\end{bmatrix} +
\begin{bmatrix}
  u_{t+1} \\
v_{t+1} \\
\eta_{t+1}
\end{bmatrix},
\tag{B11}
$$

where $E_x = (I - A_{22})^{-1}\theta$, $E_r = a + A_{12}E_x$, and without loss of generality $E_\pi = 0$. We begin working with multiple assets, so that not only $x_t$ but also $r_t$ and $\pi_t$ are vectors. We assume the errors in (B11) are i.i.d. across $t = 1, \ldots, T$:

$$
\begin{bmatrix}
  u_t \\
  v_t \\
  \eta_t
\end{bmatrix} \sim N\left(\begin{bmatrix}
  0 \\
  0 \\
  0
\end{bmatrix}, \begin{bmatrix}
  \Sigma_{uu} & \Sigma_{uv} & \Sigma_{u\eta} \\
  \Sigma_{vu} & \Sigma_{vv} & \Sigma_{v\eta} \\
  \Sigma_{u\eta} & \Sigma_{v\eta} & \Sigma_{\eta\eta}
\end{bmatrix}\right)\tag{B12}
$$

We assume the eigenvalues of both $A_{22}$ and $A_{33}$ lie inside the unit circle. As the elements of $\Sigma_{\eta\eta}$ approach zero, the above model approaches the perfect-predictor specification,

$$
\begin{bmatrix}
r_{t+1} - E_r \\
x_{t+1} - E_x
\end{bmatrix} =
\begin{bmatrix}
  0 & A_{12} \\
  0 & A_{22}
\end{bmatrix}
\begin{bmatrix}
r_t - E_r \\
x_t - E_x
\end{bmatrix} +
\begin{bmatrix}
  u_{t+1} \\
v_{t+1}
\end{bmatrix}.
\tag{B13}
$$

Let $\bar{A}$ denote the entire coefficient matrix in (B11), and let $\Sigma$ denote the entire covariance matrix in (B12). Define the vector

$$
\zeta_t = \begin{bmatrix}
r_t \\
x_t \\
\pi_t
\end{bmatrix},
\tag{B14}
$$

and let $V_{\zeta\zeta}$ denote its unconditional covariance matrix. Then

$$
V_{\zeta\zeta} =
\begin{bmatrix}
  V_{rr} & V_{rx} & V_{r\pi} \\
  V_{xr} & V_{xx} & V_{x\pi} \\
  V_{\pi r} & V_{\pi x} & V_{\pi\pi}
\end{bmatrix} = \bar{A}V_{\zeta\zeta}\bar{A}' + \Sigma,
\tag{B15}
$$
which can be solved as
\[
\text{vec} (V_{\zeta \zeta}) = [I - (A \otimes \bar{A})]^{-1} \text{vec} (\Sigma),
\]  
(B16)
using the well known identity vec\((DFG) = (G' \otimes D)\text{vec} (F)\).

Let \(z_t\) denote the vector of the observed data at time \(t\),
\[
z_t = \begin{bmatrix} r_t \\ x_t \end{bmatrix}.
\]
Denote the data we observe through time \(t\) as \(D_t = (z_1, \ldots, z_t)\), and note that our complete data consist of \(D_T\). Also define
\[
E_z = \begin{bmatrix} E_r \\ E_x \end{bmatrix}, \quad V_{zz} = \begin{bmatrix} V_{rr} & V_{rx} \\ V_{xr} & V_{xx} \end{bmatrix}, \quad V_{z\pi} = \begin{bmatrix} V_{r\pi} \\ V_{x\pi} \end{bmatrix}.
\]  
(B17)

Let \(\phi\) denote the full set of parameters in the model, \(\phi \equiv (\bar{A}, \Sigma, E_z)\), and let \(\pi\) denote the full time series of \(\pi_t\), \(t = 1, \ldots, T\). To obtain the joint posterior distribution of \(\phi\) and \(\pi\), denoted by \(p(\phi, \pi|D_T)\), we use an MCMC procedure in which we alternate between drawing \(\pi\) from the conditional posterior \(p(\pi|\phi, D_T)\) and drawing \(\phi\) from the conditional posterior \(p(\phi|\pi, D_T)\). The procedure for drawing \(\pi\) from \(p(\pi|\phi, D_T)\) is described in Section B3.1. The procedure for drawing \(\phi\) from \(p(\phi|\pi, D_T) \propto p(\phi)p(D_T, \pi|\phi)\) is described in Section B3.2.

### B3.1. Drawing \(\pi\) given the parameters

To draw the time series of the unobservable values of \(\pi_t\) conditional on the current parameter draws, we apply the forward filtering, backward sampling (FFBS) approach developed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994). See also West and Harrison (1997, chapter 15).

#### B3.1.1. Filtering

The first stage follows the standard methodology of Kalman filtering. Define
\[
a_t = \mathbb{E}(\pi_t|D_{t-1}) \quad b_t = \mathbb{E}(\pi_t|D_t) \quad e_t = \mathbb{E}(z_t|\pi_t, D_{t-1}) \\
f_t = \mathbb{E}(z_t|D_{t-1}) \quad P_t = \text{Var}(\pi_t|D_{t-1}) \quad Q_t = \text{Var}(\pi_t|D_t) \\
R_t = \text{Var}(z_t|\pi_t, D_{t-1}) \quad S_t = \text{Var}(z_t|D_{t-1}) \quad G_t = \text{Cov}(z_t, \pi_t|D_{t-1})
\]  
(B18)
(B19)
(B20)

Conditioning on \(\phi\) is assumed throughout this section but suppressed in the notation for convenience. First observe that
\[
\pi_1|D_0 \sim N(a_1, P_1),
\]  
(B21)
where $D_0$ denotes the null information set, so that the unconditional moments of $\pi_0$ are given by $a_1 = E_\pi = 0$ and $P_1 = V_{\pi\pi}$. Also,

$$z_1|D_0 \sim N(f_1, S_1),$$

where $f_1 = E_\pi$ and $S_1 = V_{\pi\pi}$. Note that

$$G_1 = V_{\pi\pi}$$

and that

$$z_1|\pi_1, D_0 \sim N(e_1, R_1),$$

where

$$e_1 = f_1 + G_1 P_1^{-1}(\pi_1 - a_1)$$

$$R_1 = S_1 - G_1 P_1^{-1}G'.$$

Combining this density with equation (B21) using Bayes rule gives

$$\pi_1|D_1 \sim N(b_1, Q_1),$$

where

$$b_1 = a_1 + P_1(P_1 + G'R_1^{-1}G)^{-1}G'R_1^{-1}(z_1 - f_1)$$

$$Q_1 = P_1(P_1 + G'R_1^{-1}G)^{-1}P_1.$$

Continuing in this fashion, we find that all conditional densities are normally distributed, and we obtain all the required moments for $t = 2, \ldots, T$:

$$a_t = A_{33}b_{t-1}$$

$$f_t = \begin{bmatrix} E_x + A_{12}(x_{t-1} - E_x) + b_{t-1} \\ E_x + A_{22}(x_{t-1} - E_x) \end{bmatrix}$$

$$S_t = \begin{bmatrix} Q_{t-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix}$$

$$G_t = \begin{bmatrix} Q_{t-1}A'_{33} \\ 0 \end{bmatrix} + \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix}$$

$$P_t = A_{33}Q_{t-1}A'_{33} + \Sigma_{\eta\eta}$$

$$e_t = f_t + G_t P_t^{-1}(\pi_t - a_t)$$

$$R_t = S_t - G_t P_t^{-1}G'.$$

$$b_t = a_t + P_1(P_t + G'R_t^{-1}G)^{-1}G'R_t^{-1}(z_t - f_t)$$

$$b_t = a_t + G'R_t^{-1}(z_t - f_t)$$

$$Q_t = P_t(P_t + G'R_t^{-1}G)^{-1}P_t.$$
The values of \( \{a_t, b_t, Q_t, S_t, G_t, P_t\} \) for \( t = 1, \ldots, T \) are retained for the next stage. Equations (B32) through (B34) are derived as

\[
\begin{bmatrix}
S_t & G_t \\
G'_t & P_t
\end{bmatrix} = \text{Var}(\zeta_t|D_{t-1}) = \bar{A}\text{Var}(\zeta_{t-1}|D_{t-1})\bar{A} + \Sigma = \bar{A} \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & Q_{t-1}
\end{bmatrix} \begin{bmatrix}
\bar{A} \\
\Sigma
\end{bmatrix} + \begin{bmatrix}
\Sigma_{uu} & \Sigma_{uv} & \Sigma_{u\eta} \\
\Sigma_{vu} & \Sigma_{vv} & \Sigma_{v\eta} \\
\Sigma_{u\eta} & \Sigma_{v\eta} & \Sigma_{\eta\eta}
\end{bmatrix}.
\]

### B3.1.2. Sampling—drawing \( \pi \)

We wish to draw \( (\pi_1, \ldots, \pi_T) \) conditional on \( D_T \) (and the parameters, \( \phi \)). The backward-sampling approach relies on the Markov property of the evolution of \( \zeta_t \) and the resulting identity,

\[
p(\zeta_1, \ldots, \zeta_T|D_T) = p(\zeta_T|D_T)p(\zeta_{T-1}|\zeta_T, D_{T-1}) \cdots p(\zeta_1|\zeta_2, D_1). \tag{B40}
\]

We first sample \( \pi_T \) from \( p(\pi_T|D_T) \), the normal density obtained in the last step of the filtering. Then, for \( t = T - 1, T - 2, \ldots, 1 \), we sample \( \pi_t \) from the conditional density \( p(\zeta_t|\zeta_{t+1}, D_t) \). (Note that the first two subvectors of \( \zeta_t \) are already observed and thus need not be sampled.) To obtain that conditional density, first note that

\[
\zeta_{t+1}|D_t \sim N \left( \begin{bmatrix}
f_{t+1} \\
a_{t+1}
\end{bmatrix}, \begin{bmatrix}
S_{t+1} & G_{t+1} \\
G'_{t+1} & P_{t+1}
\end{bmatrix} \right),
\]

\[
\zeta_t|D_t \sim N \left( \begin{bmatrix}
\rho_t \\
x_t \\
b_t
\end{bmatrix}, \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & Q_t
\end{bmatrix} \right),
\]

and

\[
\text{Cov}(\zeta_t, \zeta'_{t+1}|D_t) = \text{Var}(\zeta_t|D_t)\bar{A}' = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & Q_t
\end{bmatrix} \begin{bmatrix}
\bar{A}_{12} & \bar{A}_{22} & 0 \\
I & 0 & \bar{A}_{33}'
\end{bmatrix} = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
Q_t & 0 & Q_tA'_{33}
\end{bmatrix}. \tag{B43}
\]
Therefore,
\[\zeta_t | \zeta_{t+1}, D_t \sim N(h_t, H_t),\]  
(B44)

where
\[
h_t = E(\zeta_t | D_t) + \left[ \text{Cov}(\zeta_t, \zeta_{t+1} | D_t) \right] \left[ \text{Var}(\zeta_{t+1} | D_t) \right]^{-1} [\zeta_{t+1} - E(\zeta_{t+1} | D_t)]
\]
\[
= \begin{bmatrix} r_t \\ x_t \\ b_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t A_{33}^t \end{bmatrix} \left[ S_{t+1} & G_{t+1} \\ G_{t+1}^t & P_{t+1} \right]^{-1} \begin{bmatrix} z_{t+1} - f_{t+1} \\ \pi_{t+1} - a_{t+1} \end{bmatrix}
\]
and
\[
H_t = \text{Var}(\zeta_t | D_t) - \left[ \text{Cov}(\zeta_t, \zeta_{t+1} | D_t) \right] \left[ \text{Var}(\zeta_{t+1} | D_t) \right]^{-1} \left[ \text{Cov}(\zeta_t, \zeta_{t+1} | D_t) \right]'
\]
\[
= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_t \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ Q_t & 0 & Q_t A_{33}^t \end{bmatrix} \left[ S_{t+1} & G_{t+1} \\ G_{t+1}^t & P_{t+1} \right]^{-1} \begin{bmatrix} 0 & 0 & Q_t \\ 0 & 0 & 0 \\ 0 & 0 & A_{33} Q_t \end{bmatrix}
\]

The mean and covariance matrix of \( \pi_t \) are taken as the relevant elements of \( h_t \) and \( H_t \).

In the rest of the Appendix, we discuss the special case (implemented in the paper) in which \( r_t \) and \( \pi_t \) are scalars. The dimensions of \( A \) and \( \Sigma \) are then \((K + 2) \times (K + 2)\), where \( K \) is the number of predictors in the vector \( x_t \). In this case, \( A_{12} \) is a \( K \times 1 \) vector, which we denote as \( b \) (distinct from \( b_t \), defined previously), and \( A_{33} \) is a scalar that we denote as \( \delta \). We also denote \( A_{22} \) as simply \( A \) (distinct from \( A \), defined previously).

**B3.2. Drawing the parameters given \( \pi \)**

**B3.2.1. Prior distributions**

With \( r_t \) and \( \pi_t \) being scalars,
\[
\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{u\eta} \\ \sigma_{vu} & \Sigma_{vv} & \sigma_{v\eta} \\ \sigma_{\eta u} & \sigma_{\eta v} & \sigma_\eta^2 \end{bmatrix} = \begin{bmatrix} \Sigma_{\xi \xi} & \sigma_{\xi \eta} \\ \sigma_{\xi \eta}' & \sigma_\eta'^2 \end{bmatrix},
\]

where \( \xi' = [u, v'] \). We wish to be informative about \( \sigma_\eta'^2 \), to varying degrees, while being noninformative about \( \Sigma_{\xi \xi} \) and \( \sigma_{\xi \eta} \). This objective is accomplished by specifying a prior for \( \Sigma \) equal to the posterior that obtains when a diffuse prior for \( \Sigma \) is combined with a hypothetical sample of \( T_0 \) observations of \( \eta_t \) and \( S_0 \) observations of \( \xi_t \), where \( S_0 \leq T_0 \) and the shorter period is a subset of the longer. This posterior is given by Stambaugh (1997), except that we impose the additional restriction that the population means of \( \eta_t \) and \( \xi_t \) are zero. The posterior in Stambaugh (1997)
relies on a change of variables, which we adopt here as well. Specifically, \( c = (1/\sigma_n^2)\sigma_\xi \eta \), and 
\( \Omega = \Sigma_{\xi \xi} - \sigma_n^2 cc' \). The hypothetical \( T_0 \) observations of \( \eta_t \) produce the sample variance \( \sigma_{\eta,0}^2 \). The 
\( S_0 \) observations of \( \xi_t \) and \( \eta_t \) produce the sample covariance matrix \( \hat{\Sigma}_0 \), giving \( \hat{\sigma}_{\eta,0}^2, \hat{c}_0 \) and \( \hat{\Omega}_0 \). (All hypothetical second moments are non-central; equivalently the hypothetical sample means are zero.) With the change in variables from \( \Sigma \) to \( (\sigma_n^2, c, \Omega) \), the latter quantities have the posterior, taken here as the prior, given by

\[
p(\sigma_n^2, c, \Omega) = p(\sigma_n^2)p(c|\Omega)p(\Omega),
\]

where

\[
\sigma_n^2 \sim \frac{T_0 \sigma_{\eta,0}^2}{\chi^2_{T_0-K-1}} \tag{B45}
\]

\[
c|\Omega \sim N \left( \hat{c}_0, \frac{1}{S_0 \hat{\sigma}_{\eta,0}^2} \Omega \right) \tag{B46}
\]

\[
\Omega \sim IW(S_0 \hat{\Omega}_0, S_0 - K) \tag{B47}
\]

Be specifying different values of \( \sigma_{\eta,0}^2 \) and \( T_0 \), we vary the degree to which the prior imposes a belief that \( \sigma_n^2 \) is close to zero, where the limiting case of \( \sigma_n^2 = 0 \) corresponds to perfect predictors. The larger we make \( T_0 \), the greater is the prior information we supply about \( \sigma_n^2 \). We set \( S_0 = K+1 \), where \( K \) is the number of predictors, which makes the prior on \( c \) and \( \Omega \) essentially noninformative (as informative as a sample of only \( K+1 \) observations, where \( K = 1 \) or 3). The specification of \( \hat{\Sigma}_0 \) is thus inconsequential, and we simply set it to a scalar times the identity matrix. For the cases with imperfect predictors, we set \( T_0 = K+2 \) and then set \( \sigma_{\eta,0} \) equal to either 0.005 or 0.01 when using annual data and either 0.001 or 0.003 when using quarterly data.

We also wish to be informative about \( \delta \), the autocorrelation of \( \pi_t \). Here we specify priors for \( \delta \) identical to those specified for the autocorrelation (\( \beta \)) of the conditional mean \( \mu_t \) in the predictive system analyzed previously. Specifically, the prior for \( \delta \) is a truncated Normal distribution, with the mean and standard deviation of the corresponding non-truncated distribution denoted as \( \delta \) and

\[\begin{align*}
p(\sigma_n^2) &\propto \sigma_n^{-(T_0-K+1)} \exp \left\{ -\frac{T_0 \sigma_{\eta,0}^2}{2\sigma_n^2} \right\} \quad \Rightarrow p(\sigma_n) \propto \sigma_n^{-(T_0-K)} \exp \left\{ -\frac{T_0 \sigma_{\eta,0}^2}{2\sigma_n^2} \right\} \\
p(c|\Omega) &\propto |\Omega|^{-\frac{(K+1)}{2}} \exp \left\{ -\frac{1}{2} (c - \hat{c}_0)' \left( \frac{1}{S_0 \hat{\sigma}_{\eta,0}^2} \Omega \right)^{-1} (c - \hat{c}_0) \right\} \\
p(\Omega) &\propto |\Omega|^{-\frac{(S_0+2)}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(S_0 \hat{\Omega}_0) \Omega^{-1} \right\}
\end{align*}\]
The distribution is then truncated to satisfy the stationarity requirement $|\delta| < 1$. We set $\sigma_\delta$ to 0.25 with annual data and to 0.15 with quarterly data, and we set $\bar{\delta}$ to 0.99 in both cases.

The priors on the remaining parameters are non-informative, except for the condition required for stationary of $x_t$ that $\rho(A)$, the spectral radius of $A$, be less than 1. Specifically, define

$$\psi = \begin{bmatrix} \text{vec} \begin{bmatrix} a & \theta' \\ b & A' \end{bmatrix} \end{bmatrix}. \quad (B48)$$

Then the prior for $\psi$ is given by

$$p(\psi) \propto \exp \left\{ -\frac{1}{2} (\psi - \bar{\psi})' V_\psi^{-1} (\psi - \bar{\psi}) \right\} \times 1_S, \quad (B49)$$

where

$$\bar{\psi} = \begin{bmatrix} 0 \\ \delta \end{bmatrix}, \quad V_\psi = \begin{bmatrix} \sigma^2_\psi (K+1)^2 & 0 \\ 0 & \sigma^2_\delta \end{bmatrix}. \quad (B50)$$

$\sigma^2_\psi$ is set to a large number, $1_S$ equals 1 under the conditions that $\rho(A) < 1$ and $|\delta| < 1$, and $1_S$ equals 0 otherwise. We also assume that the priors for $\Sigma$ and $\psi$ are independent.

**B3.2.2. Posterior distributions**

Define

$$y_{t+1} = \begin{bmatrix} r_{t+1} - \pi_t \\ x_{t+1} \\ \pi_{t+1} \end{bmatrix}, \quad Y_t = \begin{bmatrix} I_{K+1} \otimes [1 \times_t'] \\ 0 \end{bmatrix}, \quad \bar{\psi}, \quad u_{t+1} = \begin{bmatrix} u_{t+1} \\ v_{t+1} \\ \eta_{t+1} \end{bmatrix},$$

$$y = \begin{bmatrix} y_2 \\ \vdots \\ y_T \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{T-1} \end{bmatrix}, \quad v = \begin{bmatrix} v_2 \\ \vdots \\ v_T \end{bmatrix}.$$ 

The sample representation of the predictive system in (B10) is then

$$y = Y \psi + v,$$

for $\psi$ defined in (B48). For tractability we employ the “conditional” likelihood, which treats values in period $t = 1$ as non-stochastic. We can then apply Gibbs sampling to draw $\psi$ and $\Sigma$. This simplification seems reasonable, given that $T = 206$ for our main results.\footnote{An alternative would be to employ the “exact” likelihood that includes the density of the values at $t = 1$ and then use the Metropolis-Hastings algorithm, with the conditional posteriors given here used as proposal densities.}
B3.2.3. Drawing $\psi$

Given the prior for $\psi$ in (B49), we can apply standard results from the multivariate regression model (e.g., Zellner, 1971) to obtain the full conditional posterior for $\psi$ as

$$\psi|\cdot \sim N(\bar{\psi}, \tilde{V}_\psi) \times 1_S$$

where

$$\bar{\psi} = \tilde{V}_\psi \left[V_\psi^{-1} \psi + Y'(I_{T-1} \otimes \Sigma^{-1})y\right]$$

and

$$\tilde{V}_\psi = \left[V_\psi^{-1} + Y'(I_{T-1} \otimes \Sigma^{-1})Y\right]^{-1}.$$  

B3.2.4. Drawing $\Sigma$

Define

$$\hat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^{T-1} (y_{t+1} - Y_{t}\psi)(y_{t+1} - Y_{t}\psi)' = \begin{bmatrix} \hat{\Sigma}_{\xi\xi} & \hat{\Sigma}_{\xi\eta} \\ \hat{\Sigma}_{\xi\eta}' & \hat{\Sigma}_{\eta} \end{bmatrix}$$

$$\bar{\Sigma} = \left[\begin{array}{c} \Sigma_{\xi\xi} \\ \Sigma_{\xi\eta} \end{array}\right] = \frac{1}{\bar{S}} \left(S_0 \hat{\Sigma}_0 + (T-1)\hat{\Sigma}\right), \bar{S} = S_0 + (T-1)$$

$$\bar{c} = (1/\hat{\sigma}_{\eta}^2)\sigma_{\xi\eta}, \bar{\Omega} = \bar{\Sigma}_{\xi\xi} - \bar{\sigma}_{\eta}^2 \bar{c} \bar{c}'$$

$$\hat{\sigma}_{\eta}^2 = \frac{1}{\bar{T}} \left(T_0 \sigma_{\eta,0}^2 + (T-1)\hat{\sigma}_{\eta}^2\right), \bar{T} = T_0 + (T-1).$$

A draw of $\Sigma$ is constructed by drawing $(\sigma_{\eta}^2, c, \Omega)$. The joint full conditional of the latter is given by

$$p(\sigma_{\eta}^2, c, \Omega|\cdot) = p(\sigma_{\eta}^2|\cdot)p(c|\Omega, \cdot)p(\Omega|\cdot),$$

where the posteriors on the right-hand side are of the same form as (B45) through (B47):

$$\sigma_{\eta}^2|\cdot \sim \chi_{T-K-1}^2$$

$$c|\Omega, \cdot \sim N\left(\bar{c}, \frac{1}{S\hat{\sigma}_{\eta}^2} \Omega\right)$$

$$\Omega|\cdot \sim IW(S\Omega, S - K).$$

The draw of $\Sigma$ is then obtained by computing $\Sigma_{\xi\xi} = \Omega + \sigma_{\eta}^2 \bar{c} \bar{c}'$ and $\sigma_{\xi\eta} = \sigma_{\eta}^2 \bar{c}$.

Our results based on 25,000 draws from the posterior distribution. First, we generate a sequence of 76,000 draws. We discard the first 1,000 draws as a “burn-in” and take every third draw from the rest to obtain a series of 25,000 draws that exhibit little serial correlation.
B3.3. Predictive variance

In addition to the notation from equations (B11), (B12), and (B14), define also

$$E_\zeta = \begin{bmatrix} E_r \\ E_x \\ 0 \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} u_t \\ v_t \\ \eta_t \end{bmatrix}, \quad (B55)$$

and

$$\zeta^e_t = \zeta_t - E_\zeta. \quad (B56)$$

Equation (B11) can then be written as

$$\zeta^e_{t+1} = \bar{A}\zeta^e_t + \epsilon_{t+1}. \quad (B57)$$

For \( i > 1 \), successive substitution using (B57) gives

$$\zeta^e_{t+i} = \bar{A}^i\zeta^e_t + \bar{A}^{i-1}\epsilon_{t+1} + \bar{A}^{i-2}\epsilon_{t+2} + \cdots + \epsilon_{t+i}. \quad (B58)$$

Define

$$\zeta_{T,T+k} = \sum_{i=1}^{k} \zeta_{T+i}, \quad (B59)$$

and

$$\zeta^e_{T,T+k} = \sum_{i=1}^{k} \zeta^e_{T+i} = \zeta_{T,T+k} - kE_\zeta. \quad (B61)$$

Summing (B57) over \( k \) periods then gives

$$\zeta^e_{T,T+k} = \left( \sum_{i=1}^{k} \bar{A}^i \right) \zeta^e_t + \left( I + \bar{A} + \cdots + \bar{A}^{k-1} \right) \epsilon_{t+1}$$

$$+ \left( I + \bar{A} + \cdots + \bar{A}^{k-2} \right) \epsilon_{t+2} + \cdots + \epsilon_{t+k}$$

$$= (\bar{A}_{k+1} - I)\zeta^e_t + \bar{A}_{k}\epsilon_{t+1} + \bar{A}_{k-1}\epsilon_{t+2} + \cdots + \epsilon_{t+k}, \quad (B59)$$

where

$$\bar{A}_i = I + \bar{A} + \cdots + \bar{A}^{i-1} \quad = (I - \bar{A})^{-1}(I - \bar{A}^i). \quad (B60)$$

It then follows that

$$E \left( \zeta^e_{T,T+k} | D_T, \phi, \pi_T \right) = (\bar{A}_{k+1} - I)\zeta^e_T, \quad (B61)$$

$$E \left( \zeta^e_{T,T+k} | D_T, \phi, \pi_T \right) = (\bar{A}_{k+1} - I)\zeta^e_T + kE_\zeta, \quad (B60)$$
and
\[
\text{Var}(\zeta_{T,T+k}|D_T, \phi, \pi_T) = \text{Var}(\zeta_{T,T+k}^e|D_T, \phi, \pi_T)
\]
\[= \sum_{i=1}^{k} \Lambda_i \Sigma \Lambda_i'. \quad (B62)
\]

The first and second moments of (B61) given \(D_T\) and \(\phi\) are given by
\[
E(\zeta_{T,T+k}|D_T, \phi) = (\bar{\Lambda}_{k+1} - I) \begin{bmatrix} r_T - E_r \\ x_T - E_x \\ b_T \end{bmatrix} + kE\zeta \quad (B63)
\]

and
\[
\text{Var}[E(\zeta_{T,T+k}|D_T, \phi, \pi_T)|D_T, \phi] = (\bar{\Lambda}_{k+1} - I) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_T \end{bmatrix} (\bar{\Lambda}_{k+1} - I)'. \quad (B64)
\]

Combining (B62) and (B64) gives
\[
\text{Var}(\zeta_{T,T+k}|D_T, \phi) = E[\text{Var}(\zeta_{T,T+k}|D_T, \phi, \pi_T)|D_T, \phi] + \text{Var}[E(\zeta_{T,T+k}|D_T, \phi, \pi_T)|D_T, \phi]
\]
\[= \sum_{i=1}^{k} \Lambda_i \Sigma \Lambda_i' + (\bar{\Lambda}_{k+1} - I) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_T \end{bmatrix} (\bar{\Lambda}_{k+1} - I)'. \quad (B65)
\]

By evaluating (B63) and (B65) for repeated draws of \(\phi\) from its posterior, the predictive variance of \(\zeta_{T,T+k}\) can be computed using the decomposition,
\[
\text{Var}(\zeta_{T,T+k}|D_T) = E\{\text{Var}(\zeta_{T,T+k}|D_T)|D_T\} + \text{Var}\{E(\zeta_{T,T+k}|D_T)|D_T\}. \quad (B66)
\]

Finally, the predictive variance of \(r_{T,T+k}\) is the (1,1) element of \(\text{Var}(\zeta_{T,T+k}|D_T)\).

**B3.4. Perfect predictors**

If the predictors are perfect, then \(\pi_t\) is absent from the first equation in (B10) and the third equation simply drops out. In that case the model consists of the two equations,
\[
\begin{align*}
r_{t+1} &= a + bx_t + u_{t+1} \quad (B67) \\
x_{t+1} &= \theta + Ax_t + v_{t+1}, \quad (B68)
\end{align*}
\]
combined with the distributional assumption on the residuals,
\[
\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{vu} & \Sigma_{vv} \end{bmatrix} \right). \quad (B69)
\]
This perfect-predictor model obtains as the limiting case of the above imperfect-predictor setting as \( \sigma^2_\eta \to 0 \). Our perfect-predictor results can be computed in that manner, by setting \( T_0 \) to a large number and \( \sigma_{\eta,0} \) to a small number. For completeness, however, we also provide here the simplified calculations that arise in that special case.

In this case \( \phi \) denotes the full set of parameters in equations (B75), (B68), and (B69), and \( \Sigma \) denotes the covariance matrix in (B69). Let \( B \) denote the matrix of coefficients in (B75) and (B68),

\[
B = \begin{bmatrix} a & \theta' \\ b & A' \end{bmatrix},
\]

and let \( \psi = \text{vec}(B) \).

**B3.4.1. Posterior distributions under perfect predictors**

We specify the prior distribution on \( \phi \) as \( p(\phi) = p(\psi)p(\Sigma) \). The prior on \( \psi \) is \( p(\psi) \propto 1_S \), where \( 1_S \) equals 1 if \( \rho(A) < 1 \) and equals 0 otherwise. The prior on \( \Sigma \) is \( p(\Sigma) \propto |\Sigma|^{-(K+2)/2} \). Define the following notation: \( r = [r_2 \ r_3 \ \cdots \ r_T]' \), \( Q^+ = [x_2 \ x_3 \ \cdots \ x_T]' \), \( Q = [x_1 \ x_2 \ \cdots \ x_{T-1}]' \), \( X = [t_{T-1} \ Q] \), where \( t_{T-1} \) denotes a \((T-1) \times 1\) vector of ones, \( Y = [r \ Q^+] \), \( \hat{B} = (X'X)^{-1}X'Y \), and \( S = (Y - X\hat{B})'(Y - X\hat{B}) \). We first draw \( \Sigma^{-1} \) from a Wishart distribution with \( T - K - 2 \) degrees of freedom and parameter matrix \( S^{-1} \). Given that draw of \( \Sigma^{-1} \), we then draw \( \psi \) from a normal distribution with mean \( \hat{\psi} = \text{vec}(\hat{B}) \) and covariance matrix \( \Sigma \otimes (X'X)^{-1} \). That draw of \( \phi \) is retained as a draw from \( p(\phi|D_T) \) if \( \rho(A) < 1 \).

**B3.4.2. Predictive variance under perfect predictors**

The conditional moments of the \( k \)-period return \( r_{T,T+k} \) are given by

\[
\begin{align*}
\mathbb{E}(r_{T,T+k}|D_T,\phi) &= ka + b'\Psi_{k-1}\theta + b'\Lambda_k x_T \\
\text{Var}(r_{T,T+k}|D_T,\phi) &= k\sigma_u^2 + 2b'\Psi_{k-1}\sigma_{vu} + b' \left( \sum_{i=1}^{k-1} \Lambda_i \Sigma_{vu} \Lambda_i' \right) b,
\end{align*}
\]

where

\[
\begin{align*}
\Lambda_i &= I + A + \cdots + A^{i-1} = (I - A)^{-1}(I - A^i) \\
\Psi_{k-1} &= \Lambda_1 + \Lambda_2 + \cdots + \Lambda_{k-1} = (I - A)^{-1}[kI - (I - A)^{-1}(I - A^k)].
\end{align*}
\]

The first term in (B71) reflects i.i.d. uncertainty. The second term reflects correlation between unexpected returns and innovations in future \( x_{T+i}'s \), which deliver innovations in future \( \mu_{T+i}'s \).
That term can be positive or negative and captures any mean reversion. The third term, always positive, reflects uncertainty about future \( x_{T+i} \)'s, and thus uncertainty about future \( \mu_{T+i} \)'s. This third term, which contains a summation, can also be written without the summation as

\[
b' \left( \sum_{i=1}^{k-1} \Lambda_i \Sigma_{vv} \Lambda_i^\prime \right) b = (b' \otimes b') \left[ (I - A)^{-1} \otimes (I - A)^{-1} \right] \left[ kI - \Lambda_k \otimes I - I \otimes \Lambda_k \right] + (I - A \otimes A)^{-1} (I - (A \otimes A)^k) \text{vec} \left( \Sigma_{vv} \right).
\]

Applying the standard variance decomposition

\[
\text{Var}(r_{T,T+k}|D_T) = \text{E}\{\text{Var}(r_{T,T+k}|D_T, \phi)|D_T\} + \text{Var}\{\text{E}(r_{T,T+k}|D_T, \phi)|D_T\}, \quad (B74)
\]

the predictive variance \( \text{Var}(r_{T,T+k}|D_T) \) can be computed as the sum of the posterior mean of the right-hand side of equation (B71) and the posterior variance of the right-hand side of equation (B70). These posterior moments are computed from the posterior draws of \( \phi \), which are described in Section B3.4.1.

**B4. Additional empirical results**

**B4.1. Predictive regressions**

This section reports the results from standard predictive regressions,

\[
r_{t+1} = a + b' x_t + \hat{e}_{t+1}, \quad (B75)
\]

for various combinations of the three predictors used in the paper. The results, obtained by OLS, are reported in Table A2. Panel A reports the results based on annual data; the results based on quarterly data are reported in Panel B. In both panels, the first three regressions contain just one predictor, while the fourth regression contains all three. The table reports the estimated coefficients as well as the \( t \)-statistics, along with the bootstrapped \( p \)-values associated with these \( t \)-statistics as well as with the \( R^2 \).

\[\text{In the bootstrap, we repeat the following procedure 20,000 times: (i) Resample } T \text{ pairs of } (\hat{v}_t, \hat{e}_t), \text{ with replacement, from the set of OLS residuals from regressions (B68) and (B75); (ii) Build up the time series of } x_t, \text{ starting from the unconditional mean and iterating forward on equation (B68), using the OLS estimates } (\hat{\theta}, \hat{A}) \text{ and the resampled values of } \hat{v}_t; (iii) Construct the time series of returns, } r_t, \text{ by adding the resampled values of } \hat{e}_t \text{ to the sample mean (i.e., under the null that returns are not predictable); (iv) Use the resulting series of } x_t \text{ and } r_t \text{ to estimate regressions (B68) and (B75) by OLS. The bootstrapped } p \text{-value associated with the reported } t \text{-statistic (or } R^2\text{) is the relative frequency with which the reported quantity is smaller than its 20,000 counterparts bootstrapped under the null of no predictability.} \]
Table A2 shows that all the predictors exhibit significant ability to predict returns, especially in the multivariate regressions that involve all three predictors. In those regressions, the estimated correlation between $e_{t+1}$ and the estimated innovation in expected return, $b'u_{t+1}$, is negative. Pástor and Stambaugh (2009) suggest this correlation as a diagnostic in predictive regressions, with a negative value being what one would hope to see for predictors able to deliver a reasonable proxy for expected return.

**B4.2. Long-horizon predictive variance**

This section provides detailed robustness results on the long-horizon predictive variance, expanding the evidence reported in the paper for both predictive systems. We examine the following cases:

- **Annual data: Baseline case reported in the paper**
  - System 1: Table A3; Figures A2, A3
  - System 2: Figure A18, left column

- **Annual data: First subperiod**
  - System 1: Table A4; Figures A4, A5
  - System 2: Figure A19, left column

- **Annual data: Second subperiod**
  - System 1: Table A5; Figures A6, A7
  - System 2: Figure A19, right column

- **Annual data: One instead of three predictors**
  - System 1: Table A6; Figures A8, A9
  - System 2: Figure A20, left column

- **Annual data: Excess instead of real returns**
  - System 1: Table A7; Figures A10, A11
  - System 2: Figure A21, left column

- **Quarterly data: Baseline case reported in the paper**
  - System 1: Table A8; Figures A12, A13
  - System 2: Figure A18, right column

- **Quarterly data: One instead of three predictors**
  - System 1: Table A9; Figures A14, A15
  - System 2: Figure A20, right column

- **Quarterly data: Excess instead of real returns**
  - System 1: Table A10; Figures A16, A17

17
System 2: Figure A21, right column

We do not report subperiod results for quarterly data because the (post-war) quarterly sample is already rather short compared to the 206-year annual sample. Parameter uncertainty generally plays a larger role in shorter samples.
Table A1
Effects of Parameter Uncertainty on 20-Year Variance Ratio

The table displays the ratio \((1/20)\text{Var}(r_{T,T+20}|D_T)/\text{Var}(r_{T+1}|D_T)\), where \(D_T\) is information used by an investor at time \(T\). The value of the ratio is computed under various parametric scenarios for \(\beta\) (autocorrelation of the conditional expected return \(\mu_t\)), \(R^2\) (fraction of variance in \(r_{t+1}\) explained by \(\mu_t\)), \(\rho_{uw}\) (correlation between unexpected returns and innovations in expected returns), \(\rho_{\mu b}\) (correlation between \(\mu_T\) and its best available estimate given \(D_T\)), and \(E_r\) (the unconditional mean return). For \(\beta\), \(R^2\), \(\rho_{uw}\), and \(\rho_{\mu b}\), each parameter is either drawn from its density in Figure A1 when uncertain or set to a fixed value. The parameters \(\beta\), \(R^2\), and \(\rho_{uw}\) are set to their medians when held fixed, while \(\rho_{\mu b}\) is fixed at its median as well as 0 and 1. The medians are 0.86 for \(\beta\), 0.12 for \(R^2\), -0.66 for \(\rho_{uw}\), and 0.70 for \(\rho_{\mu b}\). The variance of \(E_r\) given \(D_T\) is either 0 (when fixed) or 1/200 times the expected variance of one-year returns (when uncertain).

<table>
<thead>
<tr>
<th>(\beta)</th>
<th>(R^2)</th>
<th>(\rho_{uw})</th>
<th>(E_r)</th>
<th>(\rho_{\mu b}) fixed at 0</th>
<th>(\rho_{\mu b}) fixed at 0.70</th>
<th>(\rho_{\mu b}) fixed at 1</th>
<th>(\rho_{\mu b}) uncertain</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>0.95</td>
<td>0.87</td>
<td>0.77</td>
<td>0.87</td>
</tr>
<tr>
<td>F</td>
<td>U</td>
<td>F</td>
<td>F</td>
<td>1.20</td>
<td>1.06</td>
<td>0.90</td>
<td>1.06</td>
</tr>
<tr>
<td>F</td>
<td>U</td>
<td>U</td>
<td>F</td>
<td>1.05</td>
<td>0.97</td>
<td>0.87</td>
<td>0.97</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>U</td>
<td>1.02</td>
<td>0.94</td>
<td>0.84</td>
<td>0.94</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>U</td>
<td>1.05</td>
<td>0.97</td>
<td>0.88</td>
<td>0.97</td>
</tr>
<tr>
<td>U</td>
<td>U</td>
<td>U</td>
<td>F</td>
<td>1.36</td>
<td>1.22</td>
<td>1.06</td>
<td>1.22</td>
</tr>
<tr>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>1.45</td>
<td>1.32</td>
<td>1.17</td>
<td>1.32</td>
</tr>
</tbody>
</table>
Table A2
Predictive Regressions

This table summarizes the results from predictive regressions $r_t = a + b' x_{t-1} + e_t$, where $r_t$ denotes real log stock market return and $x_{t-1}$ contains the predictors (listed in the column headings) lagged by one year. Innovations in expected returns are constructed as $b' v_t$, where $v_t$ contains the disturbances estimated in a vector autoregression for the predictors, $x_t = \theta + A x_{t-1} + v_t$. The table reports the estimated slope coefficients $\hat{b}$, the correlation $\text{Corr}(e_t, b' v_t)$ between unexpected returns and innovations in expected returns, and the (unadjusted) $R^2$ from the predictive regression. The independent variables are rescaled to have unit variance. The correlations and $R^2$'s are reported in percent (i.e., $\times 100$). The OLS $t$-statistics are given in parentheses “( )”. The $t$-statistic of $\text{Corr}(e_t, b' v_t)$ is computed as the $t$-statistic of the slope from the regression of the sample residuals $\hat{e}_t$ on $\hat{b} \hat{v}_t$. The $p$-values associated with all $t$-statistics and $R^2$'s are computed by bootstrapping and reported in brackets “[ ]”. Each $p$-value is the relative frequency with which the reported quantity is smaller than its 20,000 counterparts bootstrapped under the null of no predictability. See footnote 3 of this document for more details on the bootstrapping procedure.

<table>
<thead>
<tr>
<th>Panel A. Annual data (1802–2007)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dividend Yield</td>
</tr>
<tr>
<td>Dividend Yield</td>
</tr>
<tr>
<td>(1.891)</td>
</tr>
<tr>
<td>[0.057]</td>
</tr>
<tr>
<td>Bond Yield</td>
</tr>
<tr>
<td>(0.690)</td>
</tr>
<tr>
<td>[0.236]</td>
</tr>
<tr>
<td>Term Spread</td>
</tr>
<tr>
<td>(2.129)</td>
</tr>
<tr>
<td>[0.018]</td>
</tr>
<tr>
<td>Bond Yield</td>
</tr>
<tr>
<td>(2.383)</td>
</tr>
<tr>
<td>[0.021]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Quarterly data (1952Q1–2006Q4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond Yield</td>
</tr>
<tr>
<td>Bond Yield</td>
</tr>
<tr>
<td>(3.299)</td>
</tr>
<tr>
<td>[0.001]</td>
</tr>
<tr>
<td>Dividend Yield</td>
</tr>
<tr>
<td>(1.951)</td>
</tr>
<tr>
<td>[0.091]</td>
</tr>
<tr>
<td>CAY</td>
</tr>
<tr>
<td>(3.866)</td>
</tr>
<tr>
<td>[0.000]</td>
</tr>
<tr>
<td>Bond Yield</td>
</tr>
<tr>
<td>(3.145)</td>
</tr>
<tr>
<td>[0.002]</td>
</tr>
</tbody>
</table>
Table A3 (same as Table 1 in the paper)
Variance Ratios and Components of Long-Horizon Variance

The first row of each panel reports the ratio \((1/k)\text{Var}(r_{T,T+k}|D_T)/\text{Var}(r_{T+1}|D_T)\), where \(\text{Var}(r_{T,T+k}|D_T)\) is the predictive variance of the \(k\)-year return based on 206 years of annual data for real equity returns and the three predictors over the 1802–2007 period. The second row reports \(\text{Var}(r_{T,T+k}|D_T)\), multiplied by 100. The remaining rows report the five components of \(\text{Var}(r_{T,T+k}|D_T)\), also multiplied by 100 (they add up to total variance). Panel A contains results for \(k = 25\) years, and Panel B contains results for \(k = 50\) years. Results are reported under each of three priors for \(\rho_{uw}, R^2\), and \(\beta\). As the prior for one of the parameters departs from the benchmark, the priors on the other two parameters are held at the benchmark priors. The “tight” priors, as compared to the benchmarks, are more concentrated towards \(-1\) for \(\rho_{uw}\), \(0\) for \(R^2\), and \(1\) for \(\beta\); the “loose” priors are less concentrated in those directions.

<table>
<thead>
<tr>
<th>Prior</th>
<th>(\rho_{uw})</th>
<th>(R^2)</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Panel A. Investment Horizon (k = 25) years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.30</td>
<td>1.36</td>
<td>1.26</td>
</tr>
<tr>
<td>IID Component</td>
<td>2.59</td>
<td>2.60</td>
<td>2.59</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-4.13</td>
<td>-4.01</td>
<td>-4.10</td>
</tr>
<tr>
<td>Uncertain Future (\mu)</td>
<td>2.91</td>
<td>2.86</td>
<td>2.84</td>
</tr>
<tr>
<td>Uncertain Current (\mu)</td>
<td>0.97</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>1.48</td>
<td>1.58</td>
<td>1.41</td>
</tr>
<tr>
<td>Panel B. Investment Horizon (k = 50) years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.76</td>
<td>1.82</td>
<td>1.64</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>5.14</td>
<td>5.34</td>
<td>4.79</td>
</tr>
<tr>
<td>IID Component</td>
<td>2.59</td>
<td>2.60</td>
<td>2.59</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-5.52</td>
<td>-5.36</td>
<td>-5.42</td>
</tr>
<tr>
<td>Uncertain Future (\mu)</td>
<td>5.40</td>
<td>5.31</td>
<td>5.13</td>
</tr>
<tr>
<td>Uncertain Current (\mu)</td>
<td>0.95</td>
<td>0.94</td>
<td>0.91</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>1.72</td>
<td>1.85</td>
<td>1.59</td>
</tr>
</tbody>
</table>
### Table A4

**Variance Ratios and Components of Long-Horizon Variance**

First subperiod (1802–1904)

Counterpart of Table A3

<table>
<thead>
<tr>
<th>Prior</th>
<th>( \rho_{uw} )</th>
<th>( R^2 )</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.28</td>
<td>1.35</td>
<td>1.38</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>2.86</td>
<td>3.00</td>
<td>3.05</td>
</tr>
<tr>
<td>IID Component</td>
<td>2.01</td>
<td>2.01</td>
<td>2.02</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-1.99</td>
<td>-1.71</td>
<td>-1.44</td>
</tr>
<tr>
<td>Uncertain Future ( \mu )</td>
<td>1.17</td>
<td>1.13</td>
<td>1.09</td>
</tr>
<tr>
<td>Uncertain Current ( \mu )</td>
<td>0.36</td>
<td>0.34</td>
<td>0.28</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>1.31</td>
<td>1.23</td>
<td>1.09</td>
</tr>
</tbody>
</table>

**Panel A.** Investment Horizon \( k = 25 \) years

<table>
<thead>
<tr>
<th>Prior</th>
<th>( \rho_{uw} )</th>
<th>( R^2 )</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.70</td>
<td>1.78</td>
<td>1.79</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>3.80</td>
<td>3.96</td>
<td>3.97</td>
</tr>
<tr>
<td>IID Component</td>
<td>2.01</td>
<td>2.01</td>
<td>2.02</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-2.49</td>
<td>-2.15</td>
<td>-1.81</td>
</tr>
<tr>
<td>Uncertain Future ( \mu )</td>
<td>2.04</td>
<td>2.00</td>
<td>1.91</td>
</tr>
<tr>
<td>Uncertain Current ( \mu )</td>
<td>0.40</td>
<td>0.37</td>
<td>0.31</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>1.84</td>
<td>1.73</td>
<td>1.54</td>
</tr>
</tbody>
</table>

**Panel B.** Investment Horizon \( k = 50 \) years
Table A5
Variance Ratios and Components of Long-Horizon Variance
Second subperiod (1905–2007)

Counterpart of Table A3

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uw}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.07</td>
<td>1.08</td>
<td>1.11</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>3.82</td>
<td>3.89</td>
<td>3.99</td>
</tr>
<tr>
<td>IID Component</td>
<td>3.28</td>
<td>3.29</td>
<td>3.29</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-4.60</td>
<td>-4.50</td>
<td>-4.37</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>3.02</td>
<td>3.03</td>
<td>3.00</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.86</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>1.26</td>
<td>1.25</td>
<td>1.24</td>
</tr>
</tbody>
</table>

Panel A. Investment Horizon $k = 25$ years

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uw}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.45</td>
<td>1.47</td>
<td>1.51</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>5.19</td>
<td>5.29</td>
<td>5.42</td>
</tr>
<tr>
<td>IID Component</td>
<td>3.28</td>
<td>3.29</td>
<td>3.29</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-5.98</td>
<td>-5.86</td>
<td>-5.71</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>5.48</td>
<td>5.52</td>
<td>5.50</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.93</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>1.48</td>
<td>1.45</td>
<td>1.46</td>
</tr>
</tbody>
</table>

Panel B. Investment Horizon $k = 50$ years
Table A6
Variance Ratios and Components of Long-Horizon Variance
One instead of three predictors (Dividend yield)

Counterpart of Table A3

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uu}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.08</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>3.14</td>
<td>3.06</td>
<td>3.06</td>
</tr>
<tr>
<td>IID Component</td>
<td>2.63</td>
<td>2.63</td>
<td>2.63</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>2.33</td>
<td>2.25</td>
<td>2.24</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.74</td>
<td>0.70</td>
<td>0.69</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>1.05</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Panel A. Investment Horizon $k = 25$ years

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uu}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.30</td>
<td>1.26</td>
<td>1.25</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>3.81</td>
<td>3.68</td>
<td>3.64</td>
</tr>
<tr>
<td>IID Component</td>
<td>2.63</td>
<td>2.63</td>
<td>2.63</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-4.62</td>
<td>-4.44</td>
<td>-4.42</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>3.94</td>
<td>3.75</td>
<td>3.72</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.66</td>
<td>0.63</td>
<td>0.61</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>1.21</td>
<td>1.10</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Panel B. Investment Horizon $k = 50$ years
Table A7
Variance Ratios and Components of Long-Horizon Variance
Excess instead of real returns

Counterpart of Table A3

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uw}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.13</td>
<td>1.17</td>
<td>1.26</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>2.94</td>
<td>3.04</td>
<td>3.24</td>
</tr>
<tr>
<td>IID Component</td>
<td>2.41</td>
<td>2.43</td>
<td>2.45</td>
</tr>
<tr>
<td>Mean Reversion $\mu$</td>
<td>-2.44</td>
<td>-1.78</td>
<td>-1.12</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>1.73</td>
<td>1.44</td>
<td>1.20</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.32</td>
<td>0.22</td>
<td>0.17</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>0.90</td>
<td>0.73</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Panel A. Investment Horizon $k = 25$ years

| Variance Ratio            | 1.36       | 1.39  | 1.44   | 1.60     | 1.39  | 1.35   | 1.43     | 1.39  | 1.40   |
| Predictive Variance       | 3.56       | 3.59  | 3.70   | 4.32     | 3.59  | 3.37   | 3.72     | 3.59  | 3.62   |
| IID Component             | 2.41       | 2.43  | 2.45   | 2.58     | 2.43  | 2.29   | 2.39     | 2.43  | 2.44   |
| Mean Reversion $\mu$      | -3.07      | -2.21 | -1.41  | -1.40    | -2.21 | -2.78  | -3.13    | -2.21 | -1.84 |
| Uncertain Future $\mu$    | 2.82       | 2.25  | 1.80   | 1.25     | 2.25  | 3.02   | 3.31     | 2.25  | 1.91   |
| Uncertain Current $\mu$   | 0.28       | 0.18  | 0.15   | 0.16     | 0.18  | 0.24   | 0.36     | 0.18  | 0.16   |
| Estimation Risk           | 1.11       | 0.94  | 0.72   | 1.73     | 0.94  | 0.61   | 0.79     | 0.94  | 0.95   |

Panel B. Investment Horizon $k = 50$ years
<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uw}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.73</td>
<td>1.92</td>
<td>1.95</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>1.13</td>
<td>1.25</td>
<td>1.27</td>
</tr>
<tr>
<td>IID Component</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-1.57</td>
<td>-1.14</td>
<td>-1.11</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>1.59</td>
<td>1.33</td>
<td>1.33</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.08</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>0.39</td>
<td>0.38</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Panel B. Investment Horizon $k = 50$ years

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uw}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>3.12</td>
<td>2.88</td>
<td>2.94</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>2.05</td>
<td>1.88</td>
<td>1.92</td>
</tr>
<tr>
<td>IID Component</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-1.84</td>
<td>-1.27</td>
<td>-1.24</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>2.34</td>
<td>1.67</td>
<td>1.68</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.07</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>0.83</td>
<td>0.80</td>
<td>0.81</td>
</tr>
</tbody>
</table>
### Table A9
Variance Ratios and Components of Long-Horizon Variance
Quarterly data (1952Q1–2006Q4)
One predictor instead of three (dividend yield)

Counterpart of Table A3

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uw}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Panel A. Investment Horizon $k = 25$ years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>3.54</td>
<td>2.20</td>
<td>2.30</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>2.36</td>
<td>1.50</td>
<td>1.57</td>
</tr>
<tr>
<td>IID Component</td>
<td>0.64</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-2.04</td>
<td>-0.76</td>
<td>-0.55</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>2.98</td>
<td>1.16</td>
<td>1.06</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.40</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>0.38</td>
<td>0.31</td>
<td>0.29</td>
</tr>
<tr>
<td>Panel B. Investment Horizon $k = 50$ years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>8.07</td>
<td>3.48</td>
<td>3.49</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>5.38</td>
<td>2.37</td>
<td>2.39</td>
</tr>
<tr>
<td>IID Component</td>
<td>0.64</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-2.76</td>
<td>-0.90</td>
<td>-0.65</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>6.19</td>
<td>1.80</td>
<td>1.62</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.42</td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>0.88</td>
<td>0.71</td>
<td>0.66</td>
</tr>
</tbody>
</table>
### Table A10

Variance Ratios and Components of Long-Horizon Variance
Quarterly data (1952Q1–2006Q4)
Excess instead of real returns

Counterpart of Table A3

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uw}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>1.35</td>
<td>1.88</td>
<td>1.89</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>0.86</td>
<td>1.19</td>
<td>1.19</td>
</tr>
<tr>
<td>IID Component</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-1.33</td>
<td>-1.00</td>
<td>-1.00</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>1.18</td>
<td>1.18</td>
<td>1.19</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.05</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>0.34</td>
<td>0.36</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Panel A. Investment Horizon $k = 25$ years

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\rho_{uw}$</th>
<th>$R^2$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tight</td>
<td>Bench</td>
<td>Loose</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>2.35</td>
<td>2.78</td>
<td>2.78</td>
</tr>
<tr>
<td>Predictive Variance</td>
<td>1.49</td>
<td>1.75</td>
<td>1.76</td>
</tr>
<tr>
<td>IID Component</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>Mean Reversion</td>
<td>-1.50</td>
<td>-1.10</td>
<td>-1.10</td>
</tr>
<tr>
<td>Uncertain Future $\mu$</td>
<td>1.60</td>
<td>1.45</td>
<td>1.46</td>
</tr>
<tr>
<td>Uncertain Current $\mu$</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Estimation Risk</td>
<td>0.73</td>
<td>0.76</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Panel B. Investment Horizon $k = 50$ years

28
Figure A1. Distributions for uncertain parameters The plots display the probability densities used to illustrate the effects of parameter uncertainty on long-run variance. In the $R^2$ panel, the solid line plots the density of the true $R^2$ (predictability given $\mu_T$), and the dashed line plots the implied density of the R-squared in a regression of returns on $b_T$. The dashed line incorporates the uncertainty about $\rho_{\mu b}$. 
Panel A. Predictive Variance of Stock Returns

Panel B. Components of Predictive Variance

Figure A2 (same as Figure 6 in the paper). Predictive variance of multiperiod return and its components. Panel A plots the variance of the predictive distribution of long-horizon returns, \( \text{Var}(r_{T,T+k}|D_T) \). Panel B plots the five components of the predictive variance. All quantities are divided by \( k \), the number of periods in the return horizon. The results are obtained by estimating the predictive system on annual real U.S. stock market returns in 1802-2007. Three predictors are used: the dividend yield, the bond yield, and the term spread.
Figure A3. Predictive variance for seven different priors. The figure plots the variance of the predictive distribution of long-horizon returns, $\text{Var}(r_{T,T+k}\mid D_T)$, as a function of the investment horizon $k$. The variance is divided by $k$, the number of periods in the return horizon. The results are obtained by estimating the predictive system on annual real U.S. stock market returns in 1802-2007. Three predictors are used: the dividend yield, the bond yield, and the term spread.
Figure A4. Predictive variance of multiperiod return and its components.
First subperiod (1802–1904). Counterpart of Figure A2.
Figure A5. Predictive variance for seven different priors.
First subperiod (1802–1904). Counterpart of Figure A3.
Figure A6. Predictive variance of multiperiod return and its components. Second subperiod (1905–2007). Counterpart of Figure A2.
Figure A7. Predictive variance for seven different priors.
Second subperiod (1905–2007). Counterpart of Figure A3.
Figure A8. Predictive variance of multiperiod return and its components.
One instead of three predictors (dividend yield). Counterpart of Figure A2.
Figure A9. Predictive variance for seven different priors.
One instead of three predictors (dividend yield). Counterpart of Figure A3.
Panel A. Predictive Variance of Stock Returns

Panel B. Components of Predictive Variance

Figure A10. Predictive variance of multiperiod return and its components.
Excess instead of real returns. Counterpart of Figure A2.
Figure A11. Predictive variance for seven different priors. Excess instead of real returns. Counterpart of Figure A3.
Figure A12. Predictive variance of multiperiod return and its components.
Quarterly data (1952Q1–2006Q4). Counterpart of Figure A2.
Figure A13. Predictive variance for seven different priors.
Quarterly data (1952Q1–2006Q4). Counterpart of Figure A3.
Figure A14. Predictive variance of multiperiod return and its components.
Quarterly data (1952Q1–2006Q4). One instead of three predictors (dividend yield). Counterpart of Figure A2.
Figure A15. Predictive variance for seven different priors.
Quarterly data (1952Q1–2006Q4). One instead of three predictors (dividend yield). Counterpart of Figure A3.
Figure A16. Predictive variance of multiperiod return and its components. Quarterly data (1952Q1–2006Q4). Excess instead of real returns. Counterpart of Figure A2.
Figure A17. Predictive variance for seven different priors.
Quarterly data (1952Q1–2006Q4). Excess instead of real returns. Counterpart of Figure A3.
Figure A18 (same as Figure 7 in the paper). Predictive variance and predictor imperfection. The plots display results under the predictive system (System 2) in which expected return depends on a vector of observable predictors, $x_t$, as well as a missing predictor, $\pi_t$, that obeys an AR(1) process. The top panels display prior distributions for $\sigma_\pi$, the standard deviation of $\pi_t$, under different degrees of predictor imperfection. The middle panels display the corresponding posteriors of $\Delta R^2$, the “true” $R^2$ for one-period returns minus the “observed” $R^2$ when conditioning only on $x_t$. The bottom panels display the predictive variances for the two imperfect-predictor cases as well for the case of perfect predictors ($\sigma_\pi = \Delta R^2 = 0$). The left-hand panels are based on annual data from 1802–2007 for real U.S. stock returns and three predictors: the dividend yield, the bond yield, and the term spread. The right-hand panels are based on quarterly data from 1952Q1–2006Q4 for real returns and three predictors: the dividend yield, CAY, and the bond yield.
Figure A19. Predictive variance and predictor imperfection. The plots display results under the predictive system (System 2) in which expected return depends on a vector of observable predictors, \( x_t \), as well as a missing predictor, \( \pi_t \), that obeys an AR(1) process. The top panels display prior distributions for \( \sigma_\pi \), the standard deviation of \( \pi_t \), under different degrees of predictor imperfection. The middle panels display the corresponding posteriors of \( \Delta R^2 \), the “true” \( R^2 \) for one-period returns minus the “observed” \( R^2 \) when conditioning only on \( x_t \). The bottom panels display the predictive variances for the two imperfect-predictor cases as well for the case of perfect predictors (\( \sigma_\pi = \Delta R^2 = 0 \)). The left-hand panels are based on annual data from 1802–1904 for real U.S. stock returns and three predictors: the dividend yield, the bond yield, and the term spread. The right-hand panels display the same results for the 1905–2007 period.
Figure A20. Predictive variance and predictor imperfection. The plots display results under the predictive system (System 2) in which expected return depends on a vector of observable predictors, $x_t$, as well as a missing predictor, $\pi_t$, that obeys an AR(1) process. The top panels display prior distributions for $\sigma_\pi$, the standard deviation of $\pi_t$, under different degrees of predictor imperfection. The middle panels display the corresponding posteriors of $\Delta R^2$, the “true” $R^2$ for one-period returns minus the “observed” $R^2$ when conditioning only on $x_t$. The bottom panels display the predictive variances for the two imperfect-predictor cases as well for the case of perfect predictors ($\sigma_\pi = \Delta R^2 = 0$). The left-hand panels are based on annual data from 1802–2007 for real U.S. stock returns and one predictor: the dividend yield. The right-hand panels are based on quarterly data from 1952Q1–2006Q4 for real returns and one predictor: the dividend yield.
Figure A21. Predictive variance and predictor imperfection. The plots display results under the predictive system (System 2) in which expected return depends on a vector of observable predictors, $x_t$, as well as a missing predictor, $\pi_t$, that obeys an AR(1) process. The top panels display prior distributions for $\sigma_\pi$, the standard deviation of $\pi_t$, under different degrees of predictor imperfection. The middle panels display the corresponding posteriors of $\Delta R^2$, the “true” $R^2$ for one-period returns minus the “observed” $R^2$ when conditioning only on $x_t$. The bottom panels display the predictive variances for the two imperfect-predictor cases as well for the case of perfect predictors ($\sigma_\pi = \Delta R^2 = 0$). The left-hand panels are based on annual data from 1802–2007 for excess U.S. stock returns and three predictors: the dividend yield, the bond yield, and the term spread. The right-hand panels are based on quarterly data from 1952Q1–2006Q4 for excess returns and three predictors: the dividend yield, CAY, and the bond yield.
References


West, Mike, and Jeff Harrison, 1997, *Bayesian Forecasting and Dynamic Models* (Springer-Verlag, New York, NY).